DOCUMENT RESUME

ED 047 590 FL 002 129

AUTHOR Struth, Johann F.
TITLE Accountability and Criterion Referenced Testing in Modern Foreign Language Programs.
PUB DATE 27 Feb 71
NOTE 14p.; Paper presented at the Florida Chapter Meeting of American Association of Teachers of Spanish and Portuguese, February 27, 1971, Ft. Meyers, Florida

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Achievement Tests, Aptitude Tests, Behavioral Objectives, Educational Improvement, Educational Objectives, Educational Testing, *Educational Trends, Evaluation Criteria, *Language Instruction, Language Proficiency, *Language Tests, *Modern Languages, Objective Tests, Performance Criteria, Performance Specifications, Program Evaluation, Second Language Learning, *Standardized Tests

ABSTRACT

The nature of educational accountability, seen as an organized, systematized process of self-evaluation, and a review of available standardized tests for foreign language programs are examined in this paper. Teacher accountability is seen to rely exclusively on the judgement of the teacher and administration within the school itself. The author comments on norm-referenced tests, criterion-referenced tests, Bloom's "formative-evaluation", Valette's "core-test concept", and the function of standardized tests. Differences among achievement, proficiency, placement, and aptitude tests are clarified. Reference to numerous standardized tests is made with emphasis on the Pimsleur Language Aptitude Battery and the Pimsleur Modern Foreign Language Proficiency Tests. (RL)

EDO 47590

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ACCOUNTABILITY AND CRITERION REFERENCED TESTING IN MODERN FOREIGN LANGUAGE PROGRAMS

By

Johann F. Struth
Educational Consultant
Foreign Languages
Test Department
Harcourt Brace Jovanovich, Inc.
757 Third Avenue
New York, New York 10017

A paper presented at the Florida Chapter Meeting of AATSP, February 27, 1971, Ft. Meyers, Florida

FL002 129

As you know, the term "Educational Accountability" has recently been embraced by educators, together with performance contracting, turnkey systems, and other associated educational phenomena which started this decade of educational innovations. The concept of "accountability" has been stressed as a new feature of recent efforts to improve education. This concept is easy to accept as a general idea but it is troublesome to delimit and clarify clearly. Still, there is a simple basic meaning common to the many variations of "accountability" which is fundamental to its wide acceptance. This meaning is that success or failure in education is NOT a matter of indifference. Results, especially poor results, cannot be accepted passively. Rather, every person and organization with a role to play in (foreign language) education must take inventory of the outcomes of THEIR specific activities and, as a result, expect commendation for success, accept blame for failure, and work to improve the future by heeding the lessons of the past.

It is when the general idea of accountability is made operational, and responsible individuals seek to set specific indicators of accountability, that problems arise. If accountability is to be meaningful, then the situation as it exists in accountability today must be accepted. The fact is that no numerical or quantitative index exists, or is likely to be devised, on which major reliance can be placed as an index of accountability. There is no escaping the need for human judgment.

Aha, you say now, whose judgment? His judgment, theirs, or mine? Clearly, we are not going to have an inspector general in a state department of education, complete with staff, trooping through the schools. This happens in France, Germany or behind the iron curtain, but not here. Neither can our society depend on the vagaries of the press, nor the climate of local opinion.

It may come as a shock to some of you, but the conclusion is inescapable. The source of accountability, both as to data and verdict, must be the school system itself. For day-to-day or year-to-year accountability, society must rely on the school system itself.

I am not going to say that a school system may not, on occasion, have some outside experts conduct a special survey, particularly when normal indicators of accountability such as standardized tests raise the suspicion that a serious drop has occurred in the effectiveness of an educational program of such a school or school district. However, if such outside surveys are bound to be the exception, in what then shall we rely as a rule? The answer must be that we must rely on the considered professional judgment of the school system. But how can such a system of accountability be effective, you may ask yourselves; or how can it function so that it is honestly and constructively critical and not a self-serving whitewash?

Accountability must be the outcome of an organized, systematized process of self-evaluation. It should be periodic, pervasive and enthusiastic. The general goal should be improvement. Improvement, in essence, begins and ends with the classroom teacher. Consequently, he must be included in the group of people who are organizing the accountability efforts. While many of the inputs into accountability are in action on a daily basis, there should be an organized expression of such as a formal procedure on an annual basis.

The goals which the teacher sets for the students in the class should be stated in behavioral terms, if possible. This means that each goal should specify a change in pupil performance, pupil outcomes. Despite the stress on behavioral objectives, assessment is not limited to overt

manifestations. Ratings and judgments by competent persons concerning attitude and motivations must be accepted. Goals must be set for each student in the whole range of important educational (foreign language) objectives by the teacher, with the assistance of his colleagues, supervisor and principal. The organizational basis necessary for the educational program is equally necessary for an assessment for accountability. The teacher is not a law unto himself.

Accountability puts new emphasis on the needs that have long been realized in planning the educational program. In setting a child's goal, it is necessary to consider the curriculum and syllabus. Increasingly, it will become necessary to specify learnings in criterion terms. But until criterion-referenced tests or other quantified indices are being developed, accountability must rely on assessment of process as a key to product. That is, is the teacher making careful plans to carry out those activities which experience has led us to rely upon as marking the paths to student achievement?

When you begin to analyze goals and possible outcomes for the purpose of accountability, the results will be familiar and homely rather than revolutionary and mod. The main sources of accountability are going to rest on the verities that produce the report card and the techniques of supervision for the improvement of instruction.

In the important area of student achievement, it is very natural to look to the results of standardized tests as an index of educational success. The apparent logic is alluring. Good teaching should result in pupil learning and a carefully constructed test provides dependable evidence as to whether or not any learning has taken place. Standardized tests are objective, free from teacher bias and provide a reference stan-

dard in terms of a national normative student group which gives meaning to the scores. As a result, standardized norm-referenced tests are almost universally accepted as indicators of group achievement in the key school subject areas. The norm-referenced test enables the teacher to compare a student's performance against that of other students. Whatever growth in learning has taken place can be measured by the change in scores from a pre-test to a post-test administration. Such scores usually are reported as standard scores, like the well-known 200-800 scale used by the College Board, as percentile ranks, or as stanines. The criterion-referenced test reports the student's foreign language proficiency in absolute terms. For example, Student 'A' speaks the language well enough to get around in the foreign country, or Student 'B' can handle the present tense but not the past tense in the indicative mood. Tests of this type are graded on a pass/fail or mastery/nonmastery basis as opposed to classroom tests of achievement that are normally graded on a letter grade basis.

Just as the emphasis in aptitude testing is shifting from the negative "who will succeed and who will fail" to the positive "how can the course be set up so that all students will succeed," so too is a change underway in the realm of classroom testing. The traditional quiz or unit test had to be difficult enough to provide a broad range of scores so that grades could be assigned with some degree of confidence. This practice of ranking students, either numerically or by means of letter grades, did furnish an incentive for the competition-minded student, but it had a stifling effect on the "C" and "D" student who found that success was consistently out of reach. Even when this student had reached a positive level of achievement

in a specific subject, the top third of the class had usually outdistanced him in terms of material covered and his achievement went unrecognized. Bloom states categorically that the traditional set of expectations is "the most wasteful and destructive aspect of the present educational system." The new trend in classroom teaching is toward promoting mastery for all the students.

In the area of foreign languages, the emphasis on mastery is of greatest importance. Pimsleur, Sundland and McIntyre in their study on underachievement pointed out the cumulative nature of second-language learning: of the students who get an A the first year, less than half will get an A the second year; more than half of those who get a B the first year will get a lower grade the second year. A serious problem facing foreign language teachers in the United States is the high attrition rate; roughly half the students in a first-year class go on to second year, only half of these progress to third year, etc. Unless the student really learns, unless he MASTERS the material presented in the first year, he will be unable to succeed in the second-year course.

Mastery learning is not a new idea but it has not always gone under that name. The word "MASTERY," as you know, is very common in educational parlance. It connotes having learned something well as promised in the adage, "Practice makes perfect." Mastery usually comes easily to the student when there is a very limited skill to be learned and one has the opportunity for abundant practice. Additionally, with mastery comes a feeling of pleasure and self-confidence to a student from a job well done.

In the study of human learning, educational psychologists long ago discovered two important principles:

(1) Given meaningfulness, learning is retained easily where there is abundant practice; and

(2) Meaningful learning is easily transferred.

"Meaningful," in this context, means bearing a relationship to previous learning. It also implies that the goals to be obtained are obvious. Transfer, in essence, means that one is able to use previous learning by applying it to solution of problems or to decision making.

Until a very few years ago, prevailing practices of instruction and evaluation of instruction promoted unsound effects on learners. Individual differences were, and still are, neglected in "lock-step instruction." In a sense, the instructional time was held constant while the amount of material varied. The normal curve was being overused and misused in evaluation processes. But mastery learning has different requirements. In mastery learning, the materials are held constant while the learning/study time is allowed to vary for individual students.

For Benjamin Bloom the strategy for mastery learning rates rests on the effective utilization of formative evaluation. The formative test covers a brief unit of instruction and is graded on a mastery/nonmastery basis. The level of mastery may be set quite high (control over 90 per cent of the material presented) but the student is given as many chances as he needs to attain the mastery level. If a student does not pass the formative test, his corrected test shows not only where his weaknesses are (diagnosis) but also suggests what he might do (listen to specific tapes, read a related presentation in another text, go over a few pages in the workbook, etc.) to remedy those weaknesses (prescription).

Dr. Rebecca Valette has suggested the core-test concept which would adapt formative evaluation to the area of foreign languages. All students enrolled in a given language course would be expected to master the core vocabulary and core structure plus the phonetic and morphophonemic systems; those students who assimilate the core material more rapdily would be given supplementary work in reading comprehension and listening comprehension. Since all students would be working on the same core material, group work in speaking and writing would be facilitated. In the place of traditional grades, report cards would indicate the number of units mastered. It is hoped that eventually colleges will word their foreign language entrance requirement in terms of a specified level of mastery rather than in terms of the number of hours (measured in "years") spent sitting in a language classroom. The adoption of such an entrance requirement has been frequently recommended.

It might interest this audience to learn of the use of criterion-referenced tests in a project that compared the effectiveness of three Spanish elementary school programs. This study was conducted by the California State Department of Education in 1966 with Gerald Newark and Ray Sweigert being the principal investigators. The striking and rather frightening conclusion was that students were attaining only a small percentage of the stated objectives of the three courses of study which were investigated. With respect to language testing, this study by Neward & Sweigert is of singular importance because it demon-

strates the feasibility of criterion-referenced testing within the context of a large-scale research project. It also leads us to question whether the traditional method of evaluating only a small sample of the linguistic course objectives might not obscure serious deficiencies in learning conditions and teaching.

In Volume I of "The Britannica Review of Foreign Language Education," several pages of Chapter 12, "Testing," are devoted to a discussion of classifying the aims of foreign language instruction. Dr. Rebecca Valette, who authored this chapter as well as a book entitled, Modern Language Testing-A Handbook, discusses how the objectives must be clearly stated in behavioral terms if the teacher intends to test whether or not these objectives have been obtained. This growing emphasis on terminal behavior, and by this I mean observable and verifiable changes in student behavior, has grown out of research in programmed instruction with which most of us are familiar by now.

Another section of Chapter 12 in "The Britannica Review of Foreign Language Education" deals with appropriate testing techniques to determine if the behavioral objectives have been attained. In this section you will also find a listing of presently available standard tests as well as information on criterion-referenced test item writing.

In the final portion of my talk, now that I have discoursed on both accountibility and the development of behavioral objectives in foreign language programs, I wish to throw some light on presently available standardized tests in our field. These tests must be classified into four categories and I trust that you will permit me to briefly describe each

one:

(1) Achievement Tests attempt to measure how much a person knows. They are so-called because a student has to "struggle" through a course or learning experience of some sort in order to experience/achieve a certain amount of control of the language.

(2) Proficiency Tests are the same kind of tests as achievement tests if they are thought of independent of a specific learning experience. This simply means that they are not representative of a particular textbook or teaching approach but rather that they are universal in structure and thus encompass the content of many textbooks and measure the effectiveness of a multitude of teaching approaches.

(3) Placement Tests is a description of an achievement or proficiency test when it is used to place students in a particular language class or experience. They thus can become criterion-referenced tests.

(4) Aptitude Tests are fundamentally different from the three other types just described. They are essentially designed to predict or prognosticate future success in a foreign language program and they tend to be endowed with diagnostic properties which enable the classroom teacher to plan instruction according to the needs of his incoming class of Spanish I students.

Now, let us literally throw some light on several transparencies I brought along to show you some types of standardized tests presently available to us.

1. Pimsleur Language Aptitude Battery
2. Pimsleur Spanish Proficiency Test
3. New York State Regents Examination, Spanish, Level III

(1) The Pimsleur Language Aptitude Battery is designed for use in junior and senior high schools. It can be administered in one sitting of 45-55 minutes duration or in two separate settings. This aptitude test consists of six parts and measures the following factors:

1 - School Performance in Part I
2 - Student Motivation in Part II
3 - Verbal Abilities in Parts III & IV
4 - Auditory Abilities in Parts V & VI

This test is designed to be used as a predictive instrument as well as a diagnostic device. Used as a predictive instrument, it will assist teachers and counselors in determining, in advance of instruction, how successful a student will be. It thus provides evidence on the basis of which students can be selected, screened and grouped. Used as a diagnostic instrument, this aptitude battery will assist the teacher in analyzing the learning difficulties that his students may encounter in studying Spanish or some other language. Such a diagnosis can be based upon a close study and analysis of test results on the various subtests in the battery. This study permits the teacher to explore a student's strengths and weaknesses and to suggest areas for enrichment or remedial help.

Two additonal aptitude tests should also be mentioned. They are published by the Psychological Corporation in New York and designed for senior high school/college consumption and for use in Grades 3-6, respectively.

Both are authored by John Carroll and Stanley Sapon as co-authors and are known as the Modern Language Aptitude Test. They, too, are tape oriented with regard to measuring auditory memory and sound-symbol associations.

In the area of standardized achievement or proficiency tests we find quite a few instruments that are available for use in secondary schools and colleges. They can be found in two categories, restricted and unrestricted. In the restricted column we find the College Board Achievement Tests which are administered annually in specified test centers. The Graduate Record Examination and the MLA Proficiency Tests for Teachers and Advanced Students would also be found in the restricted group in addition to the College Board Advanced Placement Tests which are administered in many high schools where such AP programs exist.

In the unrestricted category of tests we can find the following: California Common Concepts Tests, 1964, designed to measure listening comprehension in English as a second language, as well as in German, French, Spanish, on two levels. This test can be categorized as a pure listening comprehension test since it involves no printed word. Other tests would be the MLA Cooperative Foreign Language Tests, 1963 and the Pimsleur Modern Foreign Language Proficiency Tests, 1967, both of which cover two or more levels of learning experience, are norm-referenced, tape oriented, and measure the four basic communication skills. In other words, both of these tests were developed according to the latest principles of test construction which are based on linguistic understanding of language and the observations concerning the role of habit in learning a foreign language.

Contrary to the MLA Cooperative Tests, the Pimsleur Proficiency Tests use separate booklets and/or answer sheets for each of the skills as well as levels tested. This simply recognizes that individual skills do not develop nor advance simultaneously. A description of these tests would be as follows:

Test I, Listening Comprehension is made up of two parts. Part I contains 20 phonemic accuracy items, utilizing whole utterances, while in Part II the student must select the most appropriate response to a spoken stimulus. Comprehension is indicated by the student's ability to select from among four written possibilities the most appropriate rejoinder of the spoken utterance.

Test II, Speaking Proficiency, is also cued and timed by the voice on tape. All student responses are recorded on tape. In Part I of this test, pictures are used as stimulito test the ability to recall vocabulary spontaneously. In Part II the student's ability in reproducing specific sounds or sound patterns in the target language are tested. Part III tests his ability to respond orally, both appropriately and adequately, to basic stimuli in the form of questions.

As for Test III, Reading Comprehension, you will notice from these passages and questions which follow, that the following learning problems are being tested: (a) understanding of words in context; (b) ability to read correctly for literal meaning, and (c) the ability to infer ideas communicated in a passage.

Test IV, Writing Proficiency, is divided into four parts, moving from simple one word insertions in Part I to the more complex tasks of writing basic verb tenses and manipulating vocabulary and grammatical forms in a given context in Parts II and III. In Part IV,

the student's ability to write sentences or a paragraph is tested. Pictorial stimuli are employed, thus carefully controlling the vocabulary to be used and the complexity of ideas.

If time permits, let me also show you how a state mandated third year Spanish test looks which measures listening comprehension, reading and writing. (3)

An awareness of the ways in which standardized tests can be used by classroom teachers, supervisors and placement counselors should be of invaluable assistance to all foreign language teachers. This task can be accomplished only as long as he remains abreast of professional developments and innovations in his field. You, ladies and gentlemen, are demonstrating that you are, indeed, professionals by the mere fact that you are present at these meetings of the Florida Chapter of AATSP. By continuing your interest in the activities of your professional organizations, by reading about and studying the results of new developments in your fields of interest, you will be in a better position to give an accounting of your efforts to all concerned.